PROJECT REPORT:

Warmup Exercise:

S.no	Task	Using Function and Graphs
1	Show top 5 records of the dataset	head(5)
2	Show bottom 10 records of the dataset	tail(10)
3	Show no. of rows and columns	shape[0],shape[1]
4	Show no of total values(elements) in the dataset	size
5	Show the column names of the dataset	columns
6	Show the data type of the columns of the dataset	dtypes
7	Show the indexes, column's datatypes and memory usage	Info()

TASK:

sno	Task	Using function and graphs
1	any duplicate records in the dataset	duplicated()
	Remove the duplicates permanently	drop_duplicates()
2	Show the null values that are present in the dataset	isnull()
	Show the count of the null values	len(df.isnull()) df.isnull().sum()

	Create a heatmap	plt.figure(figsize=(10, 8)) sns.heatmap(df.isnull()) plt.title('Null Value Heatmap') plt.xlabel('Rows') plt.ylabel('Columns') plt.show()
3	:For 'House of Cards', what is the show id and director name	head()
4	year highest number of tv shows and movies were released	groupby value_counts() sort_index() max()
	Show the count for individual	Groupby() size() reset_index

| | | Show them in bar graph as well | ```
plt.figure(figsize=(12, 6))
sns.barplot(x='Release_Date', y='Count', data=df3)
plt.title('Netflix Releases by Year')
plt.xlabel('Release Year')
plt.ylabel('Count')
plt.show()
``` |
|---|---|---|---|
| 5 | | movies and tv shows are there in the dataset | + |
| | | Show the count of movies and tv shows separately | Len() |
| | | Show them using count plot as well | ```
plt.figure(figsize=(8, 6))
sns.countplot(x='Category', data=df)
plt.title('Movie and TV Show Count')
plt.xlabel('Type')
plt.ylabel('Count')
plt.show()
``` |

6	Show the movies that were released in the year 2000	(&)operator
	Show the title of tv shows that were released only in india	Isin()
	Show the Top 10 directors who gave the highest number of TV shows and movies to Netflix	value_counts() head(10)

		Show all the records where "category is movie and type is comedies" or "county is united kingdom"	(&)operator
		In how many movies/shows, Tom Cruise were cast	df[(((df['Cast'] == "Tom Cruise"))]
7		What are all Ratings defined by Netflix	df['Rating'].unique()
		Distribution of content ratings in a graph	plt.figure(figsize=(10, 6)) sns.countplot(x='Rating', data=df) plt.title('Distribution of Content Ratings') plt.xlabel('Rating') plt.ylabel('Count') plt.show()
		Show the ratings and no. of unique ratings	Len()

	How many movies got "TV-14" rating in Canada	(&)operator
	How many Tv Shows got rating "R" after year 2018	(&)operator
8	What is the maximum duration of a movie/tv show in Netflix?	Max()
	Show the lowest as well	Min()

9	Which country has the highest no. of .Tv Shows	value_counts()
	Show all the countries and no .of tv shows in a graph as well	plt.figure(figsize=(10, 6)) tv_shows_by_country.head(10).plot(kind='bar') plt.title('TV Shows by Country') plt.xlabel('Country') plt.ylabel('Count') plt.show()
	Compare each countries tv show and movie count and show them in a graph as well	plt.figure(figsize=(10, 6)) plt.subplot(1, 2, 1) tv_shows_by_country.head(10).plot(kind='bar') plt.title('TV Shows by Country') plt.xlabel('Country') plt.ylabel('Count')
10	Sort the dataset by year and show them	sort_values()

| | Show the count of movies/tv shows released in a year in line chart | ```
plt.figure(figsize=(10, 6))
year_count.plot(kind='line')
plt.title('Movies/TV Shows Released by Year')
plt.xlabel('Year')
plt.ylabel('Count')
plt.show()
``` |
|---|---|---|
| | Top 5 directors in the Netflix based on movies/tv shows directed | ```
value_counts()
head(5)
``` |